Global Complexity Bound of a Proximal ADMM for Linearly-Constrained Nonseperable Nonconvex Composite Programming*

Weiwei Kong† and Renato D.C. Monteiro‡

October 24, 2021

Abstract

This paper proposes and analyzes a dampened proximal alternating direction method of multipliers (DP.ADMM) for solving linearly-constrained nonconvex optimization problems where the smooth part of the objective function is nonseparable. Each iteration of DP.ADMM consists of: (ii) a sequence of partial proximal augmented Lagrangian (AL) updates, (ii) an under-relaxed Lagrange multiplier update, and (iii) a novel test to check whether the penalty parameter of the AL function should be updated. Under a basic Slater point condition and some requirements on the dampening factor and under-relaxation parameter, it is shown that DP.ADMM obtains a first-order stationary point of the constrained problem in $\mathcal{O}(\varepsilon^{-3})$ iterations for a given numerical tolerance $\varepsilon > 0$. One of the main novelties of the paper is that convergence of the method is obtained without requiring any rank assumptions on the constraint matrices.

1 Introduction

This paper presents a dampened proximal alternating direction method of multipliers (DP.ADMM) for finding approximate stationary points of the nonseparable nonconvex composite optimization problem

$$\min_{(x,y)\in\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}} \left\{\phi(x,y) := f(x,y) + h_1(x) + h_2(y) : Ax + By = d\right\}, \tag{1}$$

where A and B are linear operators, h_1 and h_2 are proper closed convex functions with compact domains, and f is a (possibly) nonconvex differentiable function on the domain of $(x,y) \mapsto h_1(x) + h_2(y)$ with a Lipschitz continuous gradient. The main idea of the method is to apply a fully proximal ADMM-type method to a sequence of penalty subproblems

$$\min_{(x,y)\in\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}} \left\{\phi(x,y) + \frac{c_\ell}{2}\|Ax + By - d\|^2\right\} \tag{2}$$

where $\{c_\ell\}_{\ell\geq 1}$ is a strictly increasing sequence of positive scalars.

*The first author has been supported by (i) the US Department of Energy (DOE) and UT-Battelle, LLC, under contract DE-AC05-00OR22725, (ii) the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration, and (iii) the IDEaS-TRIAD Fellowship (NSF Grant CCF-1740776).

†Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37830. wwkong92@gmail.com

‡School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. monteiro@isye.gatech.edu

The proximal ADMM iteration scheme for the ℓ^{th} subproblem is based on the dampened augmented Lagrangian (AL) function

$$\mathcal{L}^{\theta}_{c_\ell}(x,y;p) := \phi(x,y) + (1-\theta)\langle p, Ax+By-d\rangle + \frac{c_\ell}{2}\|Ax+By-d\|^2, \tag{3}$$

for $\theta \in (0,1)$, and it consists of the following updates at the k^{th} iteration: given $(x_{k-1}, y_{k-1}, p_{k-1})$, c_ℓ, and λ, compute

$$x_k = \underset{x\in\mathbb{R}^{n_1}}{\operatorname{argmin}} \left\{ \lambda \mathcal{L}^{\theta}_{c_\ell}(x, y_{k-1}; p_{k-1}) + \frac{1}{2}\|x - x_{k-1}\|^2 \right\}, \tag{4}$$

$$y_k = \underset{y\in\mathbb{R}^{n_2}}{\operatorname{argmin}} \left\{ \lambda \mathcal{L}^{\theta}_{c_\ell}(x_k, y; p_{k-1}) + \frac{1}{2}\|y - y_{k-1}\|^2 \right\}, \tag{5}$$

$$p_k = (1-\theta)p_{k-1} + \chi c_\ell (Ax_k + By_k - d), \tag{6}$$

where $\chi \in (0,1)$ is a suitably chosen under-relaxation parameter. Moreover, the DP.ADMM introduces a novel approach for updating c_ℓ in subproblem (2) in order to ensure fast convergence.

Under a suitable choice of the stepsize λ and the following Slater-like assumption:

$$\exists (x_\iota, y_\iota) \in \operatorname{int}(\operatorname{dom} h_1 \times \operatorname{dom} h_2) \text{ such that } Ax_\iota + By_\iota = d, \tag{7}$$

where $\operatorname{int}(\operatorname{dom} h_1 \times \operatorname{dom} h_2)$ denotes the interior of the domain of $(x,y) \mapsto h_1(x) + h_2(y)$, it is shown that for any tolerance pair $(\rho, \eta) \in \mathbb{R}^2_{++}$, the DP.ADMM obtains a pair $([\bar{x}, \bar{y}, \bar{p}], [\bar{v}^x, \bar{v}^y])$ satisfying

$$\begin{gathered} \bar{v}^x \in \nabla_x f(\bar{x},\bar{y}) + A^*\bar{p} + \partial h_1(\bar{x}), \quad \bar{v}^y \in \nabla_y f(\bar{x},\bar{y}) + B^*\bar{p} + \partial h_2(\bar{y}), \\ \|(\bar{v}^x, \bar{v}^y)\| \le \rho, \quad \|A\bar{x} + B\bar{y} - d\| \le \eta. \end{gathered} \tag{8}$$

in $\mathcal{O}(\rho^{-2}\eta^{-1})$ iterations (across all penalty subproblems).

Related Works. Since ADMM-type methods where f is convex have been well-studied in the literature (see, for example, [1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 19, 20]), we make no further mention of them here. Instead, we discuss ADMM-type methods where f is strictly nonconvex.

Denoting $\operatorname{Im}(P)$ to be the image of a linear operator P and δ_S to be the indicator function of a closed convex set (see Subsection 1.1), Table 1 presents a list of common assumptions found the literature.

Using Table 1's notation, Table 2 presents a comparison between our proposed DP.ADMM and other relevant nonconvex ADMM-type methods, under a common tolerance $\varepsilon = \min\{\rho, \eta\}$.

Before ending this review, we make some additional remarks about results in papers [14, 21] compared to the results in this paper. First, both complexity bounds in [14, 21] require that that a feasible point is readily available, while DP.ADMM does not. Second, the $\mathcal{O}(\varepsilon^{-6})$ complexity bound established in [14] is for an ADMM-type method applied to a reformulation of (1), while DP.ADMM is applied to (1) directly. Third, the method in [21] uses a small stepsize (proportional to η^2) linearized proximal gradient update while DP.ADMM considers a large stepsize (proportional to the inverse of the weak-convexity constant of f) proximal point update as in (4) and (5). Finally, it worth mentioning that [21] establishes an improved $\mathcal{O}(\varepsilon^{-3})$ complexity bound for SDD-ADMM under the additional strong assumption that $\mathcal{R}_1$ in Table 1 holds and $\prod_{t=1}^{\Gamma} \partial h(x)$ is compact for every x in the sublevel set of ϕ.

[1] See [3, 13] for a definition.

[2] Contains a proximal update for the first block and a non-proximal update in the second block.

[3] Contains a linearized proximal update instead of a fully proximal one.

$\mathcal{R}_0$	$\mathrm{Im}(A) \supseteq \{d\} \cup \mathrm{Im}(B)$.
$\mathcal{R}_1$	A has full row rank.
$\mathcal{R}_2$	A has full column rank.
$\mathcal{E}$	Problem (1) has a solution.
$\mathcal{S}$	The Slater-like assumption (7) holds.
$\mathcal{KL}$	The classical AL function, i.e. (3) with $\theta = 0$, has the KL property.[1]
$\mathcal{M}$	f is separable, i.e. $f(x^1, ..., x^\Gamma) = \sum_{t=1}^{\Gamma} f_t(x^t)$.
$\mathcal{H}$	$h_i \equiv 0$ or $h_i \equiv \delta_P$ for $i \in \{1, ..., \Gamma\}$, where P is a polyhedral set.
$\mathcal{F}$	A point $(x_0^1, ..., x_0^\Gamma) \in \prod_{t=1}^{\Gamma} \mathrm{dom}\, h_t$ satisfying $\sum_{t=1}^{\Gamma} A_t x_0^t = d$ is available.

Table 1: Common nonconvex ADMM assumptions and regularity conditions.

Algorithm	Iteration Complexity	Fully Proximal?	Assumptions
ADMM [22]	None	✗	$\mathcal{R}_2$,$\mathcal{E}$,$\mathcal{M}$
ADMM [23]	None	✗	$\mathcal{R}_0$,$\mathcal{KL}$
IAPADMM [3]	None	✗[2]	$\mathcal{R}_0$,$\mathcal{M}$,$\mathcal{H}$,$\mathcal{KL}$
IAPADMM [13]	None	✓	$\mathcal{R}_0$,$\mathcal{R}_2$,$\mathcal{M}$,$\mathcal{H}$,$\mathcal{KL}$
LPADMM [24]	None	✓	$\mathcal{H}$,$\mathcal{S}$
PADMM-m [14]	$\mathcal{O}(\varepsilon^{-6})$	✓	$\mathcal{R}_1$,$\mathcal{F}$
SDD-ADMM [21]	$\mathcal{O}(\varepsilon^{-4})$	✗[3]	$\mathcal{F}$
DP.ADMM	$\mathcal{O}(\varepsilon^{-3})$	✓	$\mathcal{S}$

Table 2: Comparison of existing nonconvex ADMM-type methods with the DP.ADMM.

Contributions. Our contributions in this paper are twofold. First, we improve the current state-of-the-art nonconvex ADMM complexity bound from $\mathcal{O}(\varepsilon^{-4})$ to $\mathcal{O}(\varepsilon^{-3})$. Second, we are the first to establish the (non-asymptotic) convergence of a nonconvex proximal ADMM-type method without relying on any restrictive assumptions on the structure of the objective function ($\mathcal{M}$, $\mathcal{KL}$, and $\mathcal{H}$ in Table 1), the properties of the constraint matrices ($\mathcal{R}_0$, $\mathcal{R}_1$, and $\mathcal{R}_2$ in Table 1), or whether or not the proposed method starts from a feasible point ($\mathcal{F}$ in Table 1).

Organization. Subsection 1.1 presents some basic definitions and notation. Section 2 presents the proposed DP.ADMM in two subsections. The first one precisely describes the problem of interest while the second one states the DP.ADMM and its iteration complexity. Section 3 presents the convergence analysis of the DP.ADMM in four subsections. The first one establishes some basic technical results, the second one presents bounds on some key algorithmic residuals, the third one bounds the Lagrange multipliers generated by the DP.ADMM, and the fourth one gives the proof of an important result, namely, Proposition 2.1. Section 4 gives some concluding remarks. Finally, the end of the paper contains several appendices.

1.1 Notation and Basic Definitions

This subsection presents notation and basic definitions used in this paper.

Let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the set of nonnegative and positive real numbers, respectively, and let $\mathbb{R}^n$ denote the n-dimensional Hilbert space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. The smallest positive singular value of a nonzero linear operator $Q : \mathbb{R}^n \to \mathbb{R}^l$ is denoted by σ_Q^+. For a given closed convex set $X \subset \mathbb{R}^n$, its boundary is denoted by ∂X and the distance of a point $x \in \mathbb{R}^n$ to X is denoted by $\operatorname{dist}_X(x)$. The indicator function of X at a point $x \in \mathbb{R}^n$ is denoted by $\delta_X(x)$ which has value 0 if $x \in X$ and $+\infty$ otherwise.

The domain of a function $h : \mathbb{R}^n \to (-\infty, \infty]$ is the set $\operatorname{dom} h := \{x \in \mathbb{R}^n : h(x) < +\infty\}$. Moreover, h is said to be proper if $\operatorname{dom} h \neq \emptyset$. The set of all lower semi-continuous proper convex functions defined in $\mathbb{R}^n$ is denoted by $\overline{\operatorname{Conv}}\, \mathbb{R}^n$. The set of functions in $\overline{\operatorname{Conv}}\, \mathbb{R}^n$ which have domain $Z \subseteq \mathbb{R}^n$ is denoted by $\overline{\operatorname{Conv}}\ Z$. The ε-subdifferential of a proper function $h : \mathbb{R}^n \to (-\infty, \infty]$ is defined by

$$\partial_\varepsilon h(z) := \{u \in \mathbb{R}^n : h(z') \geq h(z) + \langle u, z' - z \rangle - \varepsilon, \quad \forall z' \in \mathbb{R}^n\} \tag{9}$$

for every $z \in \mathbb{R}^n$. The classical subdifferential, denoted by $\partial h(\cdot)$, corresponds to $\partial_0 h(\cdot)$. The normal cone of a closed convex set C at $z \in C$, denoted by $N_C(z)$, is defined as

$$N_C(z) := \{\xi \in \mathbb{R}^n : \langle \xi, u - z \rangle \leq \varepsilon, \quad \forall u \in C\}.$$

If ψ is a real-valued function which is differentiable at $\bar{z} \in \mathbb{R}^n$, then its affine approximation $\ell_\psi(\cdot, \bar{z})$ at $\bar{z}$ is given by

$$\ell_\psi(z; \bar{z}) := \psi(\bar{z}) + \langle \nabla \psi(\bar{z}), z - \bar{z} \rangle \quad \forall z \in \mathbb{R}^n. \tag{10}$$

For conciseness, we denote Δ_k^s to be the difference of iterates for the variable s at iteration k, i.e. $\Delta_k^s = s_k - s_{k-1}$. To reduce bracketing, if $z = (x, y)$ then $f(x, y)$ is equivalent to $f(z) = f((x, y))$.

2 Alternating Direction Method of Multipliers

This section contains two subsections. The first one precisely describes the problem of interest and its underlying assumptions, while the second one presents the DP.ADMM and its corresponding iteration complexity.

2.1 Problem of Interest

This subsection presents the main problem of interest and the assumptions underlying it.

Our problem of interest is finding approximate stationary points of (1) where (ϕ, h_1, h_2) and (A, B, d) are assumed to satisfy the following assumptions:

(A1) $h_1 \in \overline{\operatorname{Conv}}\ X$ and $h_2 \in \overline{\operatorname{Conv}}\ Y$ for compact convex sets $X \subseteq \mathbb{R}^{n_1}$ and $Y \subseteq \mathbb{R}^{n_2}$;

(A2) $\phi_* := \inf_{(x,y) \in X \times Y} \phi(x, y) > -\infty$ and $\overline{\phi} := \sup_{(x,y) \in X \times Y} \phi(x, y) < \infty$;

(A3) A and B are nonzero matrices and $\mathcal{F} := \{(x, y) \in X \times Y : Ax + By = d\} \neq \emptyset$.

Moreover, it is assumed that f and $h = h_1 + h_2$ satisfy the following assumptions:

(A4) the function $h(x, y) := h_1(x) + h_2(y)$ is K_h-Lipschitz continuous on $X \times Y$ for some $K_h \geq 0$;

(A5) f is continuously differentiable on $X \times Y$ and, for every $x, x' \in X$ and $y, y' \in Y$, it holds that

$$\|\nabla_x f(x', y') - \nabla_x f(x, y)\| \leq M_{xx}\|x' - x\| + M_{xy}\|y' - y\|, \tag{11}$$
$$-\frac{m_x}{2}\|x - x'\|^2 \leq f(x', y) - \left[f(x, y) + \langle \nabla_x f(x, y), x' - x\rangle\right], \tag{12}$$

for some $(m_x, M_{xx}, M_{xy}) \in \mathbb{R}^3_+$ and

$$\|\nabla_y f(x', y') - \nabla_y f(x, y)\| \leq M_{yx}\|x' - x\| + M_{yy}\|y' - y\|, \tag{13}$$
$$-\frac{m_y}{2}\|y' - y\|^2 \leq f(x, y') - \left[f(x, y) + \langle \nabla_y f(x, y), y' - y\rangle\right], \tag{14}$$

for some $(m_y, M_{yy}, M_{yx}) \in \mathbb{R}^3_+$;

(A6) there exists $z_\iota = (x_\iota, y_\iota) \in \mathcal{F}$ such that $d_\iota := \mathrm{dist}_{\partial Z}(z_\iota) > 0$.

We now give a few remarks about the above assumptions. First, it is well known that (11) (resp. (13)) implies (12) (resp. (14)) with $m_x = \max\{M_{xx}, M_{xy}\}$ (resp. $m_y = \max\{M_{yy}, M_{yx}\}$). However, we show that better iterations complexities can be derived when scalars satisfying $m_x < \min\{M_{xx}, M_{xy}\}$ and $m_y < \{M_{yy}, M_{yx}\}$ are available. Second, condition (12) (resp. (14)) implies that $f(\cdot, y) + m_x\|\cdot\|^2/2$ (resp. $f(x, \cdot) + m_y\|\cdot\|^2/2$) is convex on X for any $y \in Y$ (resp. Y for any $x \in X$). Third, since $X \times Y$ is compact by (A1), the image of any continuous $\mathbb{R}^n$-valued function, e.g., $\nabla f(\cdot)$, is bounded.

In addition to the above remarks, we take note of the following aggregated constants:

$$\begin{aligned} &n := n_1 + n_2, \quad Q := [A \ B], \quad Z = X \times Y, \\ &D_z := \sup_{z, z' \in Z}\|z - z'\|, \quad G_f := \sup_{z \in Z}\|\nabla f(z)\|, \end{aligned} \tag{15}$$

In particular, we note the following relationships: (i) the expression $Ax + By = d$ is equivalent to $Q(x, y) = d$; (ii) D_z and G_f are finite in view of (A1) and the third remark in the previous paragraph.

We now briefly discuss the notion of an approximate stationary point of (1) in (8). It is well-known that the first-order necessary condition for a pair $(\bar{x}, \bar{y})$ to be a local minimum of (1) is that there exists a multiplier $\bar{p} \in$ such that

$$0 \in \nabla_x f(\bar{x}, \bar{y}) + A^*\bar{p} + \partial h_1(\bar{x}), \quad 0 \in \nabla_y f(\bar{x}, \bar{y}) + B^*\bar{p} + \partial h_2(\bar{y}),$$
$$A\bar{x} + B\bar{y} = d.$$

Hence, the requirements in (8) can be viewed as a direct relaxation of the above conditions. For ease of future reference, we explicitly label the problem of obtaining (8) in the problem below.

Problem $\mathcal{LCCO}$: Given a tolerance pair $(\rho, \eta) \in \mathbb{R}^2_{++}$, find a pair $([\bar{x}, \bar{y}, \bar{p}], [\bar{v}^x, \bar{v}^y])$ satisfying (8).

2.2 DP.ADMM

This subsection presents the DP.ADMM and its corresponding iteration complexity.

We state the DP.ADMM in Algorithm 2.1. Its main steps are: (i) applying the proximal updates in (4)–(5); (ii) computing several refined quantities that are related to the termination condition

Algorithm 2.1: Dampened Proximal ADMM (DP.ADMM)

Require: $(\rho, \eta) \in \mathbb{R}^2_{++}$, $(\zeta, \xi) \in \mathbb{R}^2_{++}$, $(\bar{x}_0, \bar{y}_0) \in Z$, $p_0 \in Q(\mathbb{R}^n)$, $\lambda \in (0, 1/\min\{m_x, m_y\})$, $c_1 \geq 1$, and $(\chi, \theta) \in (0,1)^2$ satisfies

$$\chi \leq \frac{\theta^2}{2(1-\theta)(2-\theta)}, \quad \theta \geq \frac{1}{2}. \tag{16}$$

Output : a pair $([\bar{x}, \bar{y}, \bar{p}], [\bar{v}^x, \bar{v}^y])$ that solves Problem $\mathcal{LCCO}$.

Function $\texttt{DP.ADMM}([f, h, A, B, d], \lambda, c_1, [\bar{x}_0, \bar{y}_0, p_0], [\zeta, \xi], [\chi, \theta], [\rho, \eta])$**:**
- **for** $\ell \leftarrow 1, 2, \ldots$ **do**
  - $\texttt{STEP 0}$ (initialize):
  - $\Phi_\ell(\cdot, \star) \leftarrow \|(1-\theta)p_0 + c_\ell(A[\cdot] + B[\star] - d)\|^2 / (2c_\ell)$
  - **find** $(x_0, y_0) \in Z$ satisfying $c_\ell \Phi_\ell(x_0, y_0) \leq \zeta$
  - **for** $k \leftarrow 1, 2, \ldots$ **do**
    - $\texttt{STEP 1}$ (prox update): ▷ Implement (4)–(5)
    - $x_k \leftarrow \mathrm{argmin}_{x \in \mathbb{R}^{n_1}} \left\{ \lambda \mathcal{L}^\theta_{c_\ell}(x, y_{k-1}; p_{k-1}) + \|x - x_{k-1}\|^2/2 \right\}$
    - $y_k \leftarrow \mathrm{argmin}_{y \in \mathbb{R}^{n_2}} \left\{ \lambda \mathcal{L}^\theta_{c_\ell}(x_k, y; p_{k-1}) + \|y - y_{k-1}\|^2/2 \right\}$
    - $\texttt{STEP 2a}$ (refinement):
    - $\varsigma_k \leftarrow \nabla_x f(x_k, y_k) - \nabla_x f(x_k, y_{k-1}) + c_\ell A^* B \Delta^y_k$
    - $(v^x_k, v^y_k) \leftarrow (-\Delta^x_k/\lambda + \varsigma_k, -\Delta^y_k/\lambda)$
    - $\tilde{p}_k \leftarrow (1-\theta)p_{k-1} + c_\ell(Ax_k + By_k - b)$
    - $\texttt{STEP 2b}$ (algorithm termination check):
    - **if** $\|(v^x_k, v^y_k)\| \leq \rho$ **and** $\|Ax_k + By_k - d\| \leq \eta$ **then**
      - **return** $([x_k, y_k, \tilde{p}_k], [v^x_k, v^y_k])$ ▷ Stop the algorithm
    - $\texttt{STEP 2c}$ (cycle termination check):
    - **if** $k \geq 2$ **then**
      - $\xi_k \leftarrow \sum_{i=2}^k c_\ell \|A^* B \Delta^y_i\| / (k-1)$
      - $r_k \leftarrow \min\limits_{2 \leq j \leq k} \{\|(v^x_j, v^y_j)\| + c_\ell^{3/2}(k-1)^{-1/2}\|Ax_j + By_j - d\|\}$
      - **if** $\xi_k \leq \xi$ **and** $r_k \leq \rho$ **then**
        - $(c_{\ell+1}, \bar{x}_\ell, \bar{y}_\ell) \leftarrow (2c_\ell, x_k, y_k)$
        - **break loop** ▷ Stop the current cycle
    - $\texttt{STEP 3}$ (multiplier update): ▷ Implement (6)
    - $p_k \leftarrow (1-\theta)p_{k-1} + \chi c_\ell(Ax_k + By_k - d)$

in Problem $\mathcal{LCCO}$; (iii) performing a novel test to determine if the method needs to be restarted with a larger penalty parameter c_ℓ; and (iv) applying the update in (6).

We now give some remarks about Algorithm 2.1. First, it is a double loop method over two sets of indices: ℓ and k. For the ease of future reference, we refer to the work performed over indices ℓ as *cycles* and the work performed indices k as *cycle iterations*. Second, for the special case of

$(\theta, \chi) = (0, 1)$, its steps 1 and 3 reduce to the classic proximal ADMM iteration

$$\begin{aligned} x_k &\leftarrow \underset{x\in\mathbb{R}^{n_1}}{\operatorname{argmin}} \left\{\lambda\mathcal{L}^0_{c_\ell}(x, y_{k-1}; p_{k-1}) + \|x - x_{k-1}\|^2/2\right\}, \\ y_k &\leftarrow \underset{y\in\mathbb{R}^{n_2}}{\operatorname{argmin}} \left\{\lambda\mathcal{L}^0_{c_\ell}(x_k, y; p_{k-1}) + \|y - y_{k-1}\|^2/2\right\}, \\ p_k &\leftarrow p_{k-1} + c_\ell(Ax_k + By_k - b). \end{aligned}$$

Consequently, the novelty of the method lies in the careful choice of (θ, χ) and the special termination condition in step 2c to ensure fast convergence of the method. Third, its step 0 can be viewed as a relaxation of the condition that the initial point of cycle is feasible, i.e. requiring that $(x_0, y_0) \in \mathcal{F}$. Finally, the computations in steps 2a–2b are performed to ensure that the its output triple aligns with the notion of a solution of Problem $\mathcal{LCCO}$.

Besides the above remarks, we briefly discuss the more (computationally) involved parts of Algorithm 2.1. First, if $(\bar{x}_0, \bar{y}_0) \in \mathcal{F}$ then $(\bar{x}_0, \bar{y}_0)$ clearly fulfills the requirements of (x_0, y_0) in step 0 with $\zeta = (1-\theta)\|p_0\|^2/2$. For the case where we only know $(x_0, y_0) \in Z$, though, we describe an efficient iterative method in Appendix B.1 which generates (x_0, y_0) from $(\bar{x}_{\ell-1}, \bar{y}_{\ell-1})$ in $\mathcal{O}(c_\ell)$ proximal, gradient, and function evaluations. Second, the computation of r_k in iteration k of step 2c appears to require both $\Theta(k)$ storage and runtime as it is evaluating the minimum of $\Theta(k)$ scalars. In Appendix B.2, we present a dynamic algorithm which only uses $\mathcal{O}(1)$ storage and runtime.

We now present the main properties of the method. To keep these statements concise, we make use of the penalty parameter-independent scalars

$$\begin{gathered} \alpha_\lambda := \min\left\{1 - \frac{\lambda m_x}{2}, 1 - \frac{\lambda m_y}{2}\right\}, \quad B_\delta := \overline{\phi} - \phi_* + \frac{1}{c_1}\left[\zeta + \frac{1}{2\chi}\|p_0\|^2\right] \\ \kappa_1 := \left(K_h + G_f + \left[M_{xy} + \frac{1}{\lambda}\right] D_z\right) D_z, \quad \kappa_2 := \frac{\chi\kappa_1 + 2\zeta c_1^{-1} + \sqrt{2\zeta}(1-\chi)(1-\theta)d_\iota\sigma_Q^+}{(\chi + \theta - \chi\theta)d_\iota\sigma_Q^+}, \\ \kappa_3 := \kappa_2 + \frac{4\chi D_z \xi}{(\chi + \theta - \chi\theta)d_\iota\sigma_Q^+}, \quad \kappa_4 := \sqrt{\frac{2B_\delta}{\chi}} + \kappa_3\sqrt{\frac{1-\theta}{c_1}}, \end{gathered} \tag{17}$$

and the penalty-dependent scalars

$$\begin{gathered} \beta_\lambda(c) := \frac{\lambda}{\alpha_\lambda}\left(\frac{1}{\lambda} + M_{xy}\right)^2 + c\|A\|^2, \quad \tau_1(c) := \max\left\{\|p_0\|, 2\left[\frac{\kappa_1 + c\|A^*B\|D_z^2}{d_\iota\sigma_Q^+}\right]\right\}, \\ \tau_2(c) := \frac{12\beta_\lambda(c)}{\chi}, \quad \tau_3(c) := \tau_2(c)\left[\frac{2}{\chi}B_\delta + \frac{(1-\theta)\tau_1^2(c)}{c}\right], \\ \tau_4(c) := \frac{1}{2}\left[\kappa_4\sqrt{\tau_2(c)} + \frac{2\sqrt{c}}{\chi}(\zeta + \kappa_3)\right] \end{gathered} \tag{18}$$

for any penalty parameter $c \geq 0$. The first result, whose proof is the topic of Section 3, presents some basic properties of the method.

Proposition 2.1. *The following statements hold:*

(a) every pair $([\bar{x}, \bar{y}, \bar{p}], [\bar{v}^x, \bar{v}^y])$ that is output by the DP.ADMM solves Problem $\mathcal{LCCO}$.

(b) the number of cycle iterations performed during the ℓ^{th} cycle is at most

$$\overline{T}_\ell(\rho) := \left\lceil 1 + \frac{\tau_3(c_\ell)}{\xi^2} + \frac{\tau_4^2(c_\ell)}{\rho^2}\right\rceil; \tag{19}$$

(c) if at the ℓ^{th} cycle we have

$$c_\ell \geq \frac{1}{2\eta}\left[\kappa_4\sqrt{\tau_2(1)} + \frac{2}{\chi}(\zeta+\kappa_3)\right] =: \overline{c}(\eta), \tag{20}$$

then the DP.ADMM stops during this cycle at step 2b.

We now establish the iteration complexity of the DP.ADMM.

Theorem 2.2. *The DP.ADMM stops and outputs a pair $([\bar{x},\bar{y},\bar{p}],[\bar{v}^x,\bar{v}^y])$ that solves Problem $\mathcal{LCCO}$ in at most*

$$\overline{T}(\eta,\rho) := \sum_{\ell=1}^{\bar{\ell}(\eta)} \overline{T}_\ell(\rho) \tag{21}$$

cycle iterations, where $\overline{\ell}(\eta) := \max\{1, \lceil \log_2[\overline{c}(\eta)/c_1])\rceil\}$. Moreover, if it holds that

$$\max\left\{\frac{1}{\xi},\frac{1}{\chi},\frac{1}{\lambda},\frac{1}{\alpha_\lambda},c_1,\zeta,\|p_0\|\right\} = \mathcal{O}(1), \tag{22}$$

then $\overline{T}(\eta,\rho) = \mathcal{O}(\rho^{-2}\eta^{-1})$.

Proof. It follows from Lemma 2.1(c) and the penalty parameter update in step 2c of the DP.ADMM that the number of cycles performed by the method is $\overline{\ell}(\eta)$. Hence, it follows from Lemma 2.1(a)–(b) and the previous observation that the DP.ADMM stops and outputs a pair that solves Problem $\mathcal{LCCO}$ in $\overline{T}(\eta,\rho) = \sum_{\ell=1}^{\bar{\ell}(\eta)} \overline{T}_\ell(\rho)$ cycle iterations. Suppose now that (22) holds. Using (22) and (18) it follows that $\tau_3(c_\ell) = \Theta(c_\ell)$ and $\tau_4(c_\ell) = \Theta(c_\ell^{1/2})$, which implies $\overline{T}_\ell(\rho) = \Theta(c_\ell\rho^{-2})$. Hence, in view of (22), the definition of $\overline{c}(\eta)$ in (20), and the fact that $c_\ell = c_1 2^{\ell-1}$ for every $\ell \geq 1$, we conclude that

$$\overline{T}(\eta,\rho) = \mathcal{O}\left(\frac{1}{\rho^2}\sum_{\ell=1}^{\overline{\ell}(\eta)} c_\ell\right) = \mathcal{O}\left(\frac{c_1}{\rho^2}\sum_{\ell=1}^{\overline{\ell}(\eta)} 2^{\ell-1}\right) = \mathcal{O}\left(\frac{\overline{c}(\eta)}{\rho^2}\right) = \mathcal{O}\left(\frac{1}{\rho^2\eta}\right).$$

□

3 Convergence Analysis

This section establishes the key properties of the DP.ADMM and contains four subsections. The first one presents some basic technical results about the iterates generated by the method, the second one presents bounds on some key residuals, the third one establishes several specialized Lagrange multiplier bounds, and the the last one gives the proof of Proposition 2.1.

Unless otherwise specified, we will focus our attention on the iterates generated during a fixed cycle $\ell \geq 1$. That is, we will analyze the cycle iterates

$$\{(x_i,y_i,p_i,\tilde{p}_i,v_i^x,v_i^y)\}_{i=1}^k, \quad \{\xi_j,r_j\}_{j=2}^k$$

generated by the DP.ADMM up to and including the k^{th} cycle iteration for $k \geq 2$. Moreover, for every $i \geq 1$, $j \geq 2$, and any $(\chi,\theta) \in \mathbb{R}^2_{++}$, we will make use of the following useful constants and shorthand notation

$$\begin{gathered} a_\theta = \theta(1-\theta), \quad b_\theta := (2-\theta)(1-\theta), \quad \alpha_{\chi,\theta} := \frac{(1-2\chi b_\theta)-(1-\theta)^2}{2\chi}, \\ z_i := (x_i,y_i), \quad v_j := (v_j^x, v_j^y), \quad f_i := Qz_i - d, \end{gathered} \tag{23}$$

as well as the averaged quantities

$$S_{p,j} := \frac{\sum_{i=2}^{j} \|p_i\|}{j-1}, \quad S_{f,j} := \frac{\sum_{i=2}^{j} \|f_i\|}{j-1}, \quad S_{v,j} := \frac{\sum_{i=2}^{j} \|v_i\|}{j-1}. \tag{24}$$

3.1 Basic Technical Results

This section presents several technical results that will be important in later analyses.

The first result presents some key relationships between the iterates generated by the method.

Lemma 3.1. *The following statements hold for every $i \le k$:*

(a) $f_i = [p_i - (1-\theta)p_{i-1}]/(\chi c_\ell)$;

(b) if $i \ge 2$, then $\chi c_\ell (f_i - f_{i-1}) = \Delta_i^p - (1-\theta)\Delta_{i-1}^p$;

(c) $v_i \in \nabla f(z_i) + Q^* \tilde{p}_i + \partial h(z_i)$ *and*

$$\|v_i\| \le \frac{1}{\lambda}\|\Delta_i^x\| + \left(\frac{1}{\lambda} + M_{xy}\right)\|\Delta_i^y\| + c_\ell \|A^* B \Delta_i^y\|.$$

Proof. (a) This is immediate from step 3 of the DP.ADMM and the definition of f_i in (23).

(b) This follows immediately from applying part (a) at indices i and $i-1$.

(c) We first prove the required inclusion. The optimality of x_k in step 1 of the DP.ADMM, assumption (A5), and the fact that $\lambda < 1/m$, first implies that

$$\begin{aligned} 0 &\in \partial \left[\mathcal{L}_{c_\ell}^\theta(\cdot, y_{i-1}; p_{i-1}) + \frac{1}{2\lambda}\|\cdot - x_{i-1}\|^2\right](x_i) \\ &= \nabla_x f(x_i, y_{i-1}) + A^* \left[(1-\theta)p_{i-1} + c_\ell(Ax_i + By_{i-1} - d)\right] + \partial h_1(x_i) + \frac{1}{\lambda}\Delta_i^x \\ &= \nabla_x f(x_i, y_i) + A^* \tilde{p}_i + \partial h_1(x_i) - \varsigma_i + \frac{1}{\lambda}\Delta_i^x \\ &= \nabla_x f(z_i) + A^* \tilde{p}_i + \partial h_1(x_i) - v_i^x. \end{aligned} \tag{25}$$

Similarly, the optimality of x_k in step 1 of the DP.ADMM, assumption (A5), and the fact that $\lambda < 1/m$, implies that

$$\begin{aligned} 0 &\in \partial \left[\mathcal{L}_{c_\ell}^\theta(x_i, \cdot; p_{i-1}) + \frac{1}{2\lambda}\|\cdot - y_{i-1}\|^2\right](y_i) \\ &= \nabla_y f(z_i) + B^* \left[(1-\theta)p_{i-1} + c_\ell(Ax_i + By_i - d)\right] + \partial h_2(y_i) + \frac{1}{\lambda}\Delta_i^y \\ &= \nabla_y f(z_i) + B^* \tilde{p}_i + \partial h_2(y_i) - v_i^y. \end{aligned} \tag{26}$$

Combining (25), (26), and the definitions of v_i, h, and Q in (23) yields the desired conclusion. To show the required inequality, we use the triangle inequality, the definition of v_i, and assumption (A5) to obtain

$$\begin{aligned} \|v_i\| &\le \frac{1}{\lambda}\left(\|\Delta_i^x\| + \|\Delta_i^y\|\right) + \|\varsigma_i\| \le \frac{1}{\lambda}\left(\|\Delta_i^x\| + \|\Delta_i^y\|\right) + \|\nabla_x f(x_i, y_i) - \nabla_x f(x_i, y_{i-1})\| + c_\ell \|A^* B \Delta_i^y\| \\ &\le \frac{1}{\lambda}\|\Delta_k^x\| + \left(\frac{1}{\lambda} + M_{xy}\right)\|\Delta_k^y\| + c_\ell \|A^* B \Delta_k^y\|. \end{aligned}$$

□

The next result presents some general bounds given by the updates in (4) and (5).

Lemma 3.2. *The following statements hold for every $i \le k$:*

(a) for every $x \in X$, it holds that

$$\lambda\left[\mathcal{L}_{c_\ell}^{\theta}(x, y_{i-1}; p_{i-1}) - \mathcal{L}_{c_\ell}^{\theta}(x_i, y_{i-1}; p_{i-1})\right] + \frac{1}{2}\|x - x_{i-1}\|^2$$
$$\ge \frac{1}{2}\|\Delta_i^x\|^2 + \left(\frac{1-\lambda m_x}{2}\right)\|x - x_i\|^2 + \frac{\lambda c_\ell}{2}\|A(x - x_i)\|^2;$$

(b) for every $y \in Y$, it holds that

$$\lambda\left[\mathcal{L}_{c_\ell}^{\theta}(x_i, y; p_{i-1}) - \mathcal{L}_{c_\ell}^{\theta}(x_i, y_i; p_{i-1})\right] + \frac{1}{2}\|y - y_{i-1}\|^2$$
$$\ge \frac{1}{2}\|\Delta_i^y\|^2 + \left(\frac{1-\lambda m_y}{2}\right)\|y - y_i\|^2 + \frac{\lambda c_\ell}{2}\|B(y - y_i)\|^2.$$

Proof. (a) Let $i \le k$ be fixed and define $\mu := 1 - \lambda m_x$ and $\|\cdot\|_\alpha^2 := \langle \cdot, (\mu I + \lambda c_\ell A^* A)(\cdot)\rangle$. Using the optimality of x_k and the fact that $\lambda \mathcal{L}_{c_\ell}^{\theta}(\cdot, y_{i-1}; p_{i-1}) + \|\cdot\|^2/2$ is μ-strongly convex with respect to $\|\cdot\|_\alpha^2$, it holds that

$$0 \in \partial\left[\lambda \mathcal{L}_{c_\ell}^{\theta}(\cdot, y_{i-1}; p_{i-1}) + \frac{1}{2}\|\cdot - x_{i-1}\|^2 - \frac{\mu}{2}\|\cdot - x_i\|_\alpha^2\right](x_i),$$

or equivalently,

$$\begin{aligned}
&\lambda \mathcal{L}_{c_\ell}^{\theta}(x_i, y_{i-1}; p_{i-1}) + \frac{1}{2}\|\Delta_i^x\|^2 \\
&\le \lambda \mathcal{L}_{c_\ell}^{\theta}(x, y_{i-1}; p_{i-1}) + \frac{1}{2}\|x - x_{i-1}\|^2 - \frac{1}{2}\|x - x_i\|_\alpha^2 \\
&= \lambda \mathcal{L}_{c_\ell}^{\theta}(x, y_{i-1}; p_{i-1}) + \frac{1}{2}\|x - x_{i-1}\|^2 - \frac{\mu}{2}\|x - x_i\|^2 - \frac{\lambda c_\ell}{2}\|A(x - x_i)\|^2,
\end{aligned}$$

which clearly implies the desired bound.

(b) The argument follows similarly as in part (a) but with $\mu = 1 - \lambda m_x$, $\|\cdot\|_\alpha^2 = \langle \cdot, (\mu_y I + \lambda c_\ell B^* B)(\cdot)\rangle$, and $\mathcal{L}_{c_\ell}^{\theta}(\cdot, y_{i-1}; p_{i-1})$ replaced by $\mathcal{L}_{c_\ell}^{\theta}(x_i, \cdot; p_{i-1})$. $\square$

The result below specializes to the previous result, and another one from Appendix A, to obtain more refined bounds.

Lemma 3.3. *The following statements hold for every $i \le k$:*

(a) $\mathcal{L}_{c_\ell}^{\theta}(z_i; p_i) - \mathcal{L}_{c_\ell}^{\theta}(z_i; p_{i-1}) = b_\theta \|\Delta_i^p\|^2/(2\chi c_\ell) + a_\theta(\|p_i\|^2 - \|p_{i-1}\|^2)/(2\chi c_\ell)$;

(b) $\mathcal{L}_{c_\ell}^{\theta}(z_i; p_{i-1}) - \mathcal{L}_{c_\ell}^{\theta}(z_{i-1}; p_{i-1}) \le -\alpha_\lambda \|\Delta_i^z\|^2/\lambda - c_\ell(\|A\Delta_i^x\|^2 + \|B\Delta_i^y\|^2)/2$;

(c) if $i \ge 2$, then it holds that

$$\frac{b_\theta}{2\chi c_\ell}\|\Delta_i^p\|^2 - \frac{c_\ell}{4}\left(\|A\Delta_i^x\|^2 + \|B\Delta_i^y\|^2\right) \le \frac{\alpha_{\chi,\theta}}{8\chi c_\ell}\left(\|\Delta_{i-1}^p\|^2 - \|\Delta_i^p\|^2\right). \tag{27}$$

Proof. (a) Using the definition of $\mathcal{L}^\theta_{c_\ell}(\cdot;\cdot)$ and Proposition 3.1(a), it holds that

$$\begin{aligned}
\mathcal{L}^\theta_{c_\ell}(z_i;p_i) - \mathcal{L}^\theta_{c_\ell}(z_i;p_{i-1}) = (1-\theta)\left\langle \Delta_i^p, f_i\right\rangle &= \left(\frac{1-\theta}{\chi c_\ell}\right)\|\Delta_i^p\|^2 + \frac{a_\theta}{\chi c_\ell}\left\langle \Delta_i^p, p_{i-1}\right\rangle \\
&= \left(\frac{1-\theta}{\chi c_\ell}\right)\|\Delta_i^p\|^2 + \frac{a_\theta}{\chi c_\ell}\left(\langle p_i, p_{i-1}\rangle - \|p_{i-1}\|^2\right) \\
&= \left(\frac{1-\theta}{\chi c_\ell}\right)\|\Delta_i^p\|^2 + \frac{a_\theta}{\chi c_\ell}\left(\frac{1}{2}\|p_i\|^2 - \frac{1}{2}\|\Delta_i^p\|^2 - \frac{1}{2}\|p_{i-1}\|^2\right) \\
&= \frac{b_\theta}{2\chi c_\ell}\|\Delta_i^p\|^2 + \frac{a_\theta}{2\chi c_\ell}\left(\|p_i\|^2 - \|p_{i-1}\|^2\right),
\end{aligned}$$

which clearly implies the desired bound.

(b) Using Lemma 3.2 with $(x,y) = (x_{i-1}, y_{i-1})$ and part (a), it holds that

$$\begin{aligned}
&\alpha_\lambda\|\Delta_i^z\|^2 + \frac{\lambda c_\ell}{2}\|A\Delta_i^x\|^2 + \frac{\lambda c_\ell}{2}\|B\Delta_i^y\|^2 \\
&\le \left(1 - \frac{\lambda m_x}{2}\right)\|\Delta_i^x\|^2 + \left(1 - \frac{\lambda m_y}{2}\right)\|\Delta_i^y\|^2 + \frac{\lambda c_\ell}{2}\|A\Delta_i^x\|^2 + \frac{\lambda c_\ell}{2}\|B\Delta_i^y\|^2 \\
&\le \lambda\left[\mathcal{L}^\theta_{c_\ell}(z_{i-1};p_{i-1}) - \mathcal{L}^\theta_{c_\ell}(z_i;p_{i-1})\right],
\end{aligned}$$

which clearly implies the desired bound.

(c) We first use Lemma A.2 with $(a,b,\tau) = (\Delta_k^p, \Delta_{k-1}^p, 4\chi b_\theta)$ to observe that

$$\|\Delta_k^p - (1-\theta)\Delta_{k-1}^p\|^2 \ge 4\chi b_\theta\|\Delta_k^p\|^2 + \chi\alpha_{\chi,\theta}\left(\|\Delta_k^p\|^2 - \|\Delta_{k-1}^p\|^2\right). \tag{28}$$

Using (28) and the inequality $(a+b)^2 \le 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}$, we then have that

$$\begin{aligned}
\frac{c}{4}\left(\|A\Delta_k^x\|^2 + \|B\Delta_k^y\|^2\right) \ge \frac{c}{8}\|Q\Delta_k^z\|^2 &= \frac{1}{8\chi^2 c}\|\Delta_k^p - (1-\theta)\Delta_{k-1}^p\|^2 \\
&\ge \frac{1}{8\chi c}\left[4b_\theta\|\Delta_k^p\|^2 + \alpha_{\chi,\theta}\left(\|\Delta_k^p\|^2 - \|\Delta_{k-1}^p\|^2\right)\right] \\
&= \frac{b_\theta}{2\chi c}\|\Delta_k^p\|^2 + \frac{\alpha_{\chi,\theta}}{8\chi c}\left(\|\Delta_k^p\|^2 - \|\Delta_{k-1}^p\|^2\right).
\end{aligned}$$

Rearranging terms, yields (27). □

3.2 Bounding Key Residuals

This subsection presents bounds on some key residuals that are important in establishing the convergence of the DP.ADMM.

We first show a unified bound on the sum of the residuals $\|\Delta_i^x\|^2$, $\|\Delta_i^y\|^2$, $c_\ell\|B\Delta_i^y\|^2$, and $c_\ell\|A\Delta_i^x\|^2$.

Lemma 3.4. *If $k \ge 2$, then it holds that*

$$\begin{aligned}
&\frac{\alpha_\lambda}{\lambda}\sum_{i=2}^k\left(\|\Delta_i^x\|^2 + \|\Delta_i^y\|^2\right) + \frac{c_\ell}{4}\sum_{i=2}^k\left(\|A\Delta_i^x\|^2 + \|B\Delta_i^y\|^2\right) \\
&\le \mathcal{L}^\theta_{c_\ell}(z_1;p_1) - \mathcal{L}^\theta_{c_\ell}(z_k;p_k) + \frac{1}{16\chi^2 c_\ell}\|\Delta_1^p\|^2 + \frac{a_\theta}{2\chi c_\ell}\|p_k\|^2 =: \delta_k.
\end{aligned}$$

Proof. Define the function

$$\Psi_i^\theta := \mathcal{L}_{c_\ell}^\theta(z_i; p_i) + \frac{\alpha_{\chi,\theta}}{8\chi c_\ell}\|\Delta_i^p\|^2 - \frac{a_\theta}{2\chi c_\ell}\|p_i\|^2 \quad \forall i \geq 1. \tag{29}$$

It follows from Lemma 3.3(b)–(c) and the definition of Ψ_i^θ above that

$$\frac{\alpha_\lambda}{\lambda}\left(\|\Delta_i^x\|^2 + \|\Delta_i^y\|^2\right) + \frac{c_\ell}{4}\left(\|A\Delta_i^x\|^2 + \|B\Delta_i^y\|^2\right) \leq \Psi_{i-1}^\theta - \Psi_i^\theta.$$

Summing the above inequality from $i = 2$ to k thus yields

$$\frac{\alpha_\lambda}{\lambda}\sum_{i=2}^{k}\left(\|\Delta_k^x\|^2 + \|\Delta_k^y\|^2\right) + \frac{c_\ell}{4}\sum_{i=1}^{k}\left(\|A\Delta_k^x\|^2 + \|B\Delta_k^y\|^2\right) \leq \Psi_1^\theta - \Psi_k^\theta. \tag{30}$$

Now, using the fact that $\alpha_{\chi,\theta} \in (0, 1/[2\chi])$, we have that

$$\Psi_1^\theta - \Psi_k^\theta \leq \mathcal{L}_{c_\ell}^\theta(z_1; p_1) - \mathcal{L}_{c_\ell}^\theta(z_k; p_k) + \frac{\alpha_{\chi,\theta}}{8\chi c_\ell}\|\Delta_1^p\|^2 + \frac{a_\theta}{2\chi c_\ell}\|p_k\|^2 \leq \delta_k. \tag{31}$$

Combining (30) and (31) yields the desired bound. □

We next bound some quantities in the potential function Ψ_j^θ from the previous result.

Lemma 3.5. *The following statements hold:*

(a) $\|p_1\|^2 \leq 2\zeta$;

(b) $\mathcal{L}_{c_\ell}^\theta(z_1; p_1) \leq \phi(z_0) + \zeta/c_\ell + a_\theta\|p_0\|^2/(2\chi c_\ell)$;

(c) $\mathcal{L}_{c_\ell}^\theta(z_k; p_k) \geq \phi(z_k) - (1-\theta)^2\|p_k\|^2/(2c_\ell)$.

Proof. (a) Using steps 0 and 3 of the DP.ADMM and Lemma B.2(b), it follows that

$$\begin{aligned}
\frac{1}{2\chi c_\ell}\|p_1\|^2 &= \frac{1}{2\chi c_\ell}\|(1-\theta)p_0 + \chi c_\ell(Qz_0 - d)\|^2 \\
&= \frac{(1-\theta)^2}{2\chi c_\ell}\|p_0\|^2 + (1-\theta)\langle p_0, Qz_0 - d\rangle + \frac{\chi c_\ell}{2}\|Qz_0 - d\|^2 \\
&\leq \frac{1}{\chi}\Phi_\ell(z_0) \leq \frac{\zeta}{\chi c_\ell},
\end{aligned}$$

which clearly implies the desired bound.

(b) Using step 0 of the DP.ADMM, Lemma B.2(b), Lemma 3.3(a)–(b) at $i = 1$, and the fact that $(a_\theta, b_\theta) \in \mathbb{R}_{++}^2$ and $\theta \in (0, 1)$, we have that

$$\begin{aligned}
\mathcal{L}_{c_\ell}^\theta(z_1; p_1) &= \mathcal{L}_{c_\ell}^\theta(z_1; p_0) - \frac{b_\theta}{2\chi c_\ell}\|\Delta_1^p\|^2 - \frac{a_\theta}{2\chi c_\ell}\left[\|p_1\|^2 - \|p_0\|^2\right] \\
&\leq \mathcal{L}_{c_\ell}^\theta(z_0; p_0) + \frac{a_\theta}{2\chi c_\ell}\|p_0\|^2 = \phi(z_0) + \Phi_\ell(z_0) + \frac{a_\theta - \chi(1-\theta)^2}{2\chi c_\ell}\|p_0\|^2 \\
&\leq \phi(z_0) + \frac{\zeta}{c_\ell} + \frac{a_\theta}{2\chi c_\ell}\|p_0\|^2.
\end{aligned}$$

(c) It holds that

$$\begin{aligned}\mathcal{L}^{\theta}_{c_\ell}(z_k;p_k) &= \phi(z_k) + (1-\theta)\left\langle p_k, Qz_k - d\right\rangle + \frac{c_\ell}{2}\|Qz_k-d\|^2 \\ &= \phi(z_k) + \frac{1}{2}\left\|\left(\frac{1-\theta}{\sqrt{c_\ell}}\right)p_k + \sqrt{c_\ell}(Qz_k-d)\right\|^2 - \frac{(1-\theta)^2}{2c_\ell}\|p_k\|^2 \\ &\geq \phi(z_k) - \frac{(1-\theta)^2}{2c_\ell}\|p_k\|^2.\end{aligned}$$

□

We now present some refined bounds on the residuals ξ_k and $S_{v,k}$.

Lemma 3.6. *Let δ_k, B_δ, and $\beta_\lambda(\cdot)$ be as in Lemma 3.4, (17), and (18), respectively. Then, it holds that*

$$\chi\delta_k \leq 2B_\delta + \frac{1-\theta}{c_\ell}\|p_k\|^2, \quad \max\{\xi_k, S_{v,k}\} \leq \sqrt{\frac{12\beta_\lambda(c_\ell)\cdot\delta_k}{k-1}}. \tag{32}$$

Proof. Using Lemma 3.5(a)–(c), the bound $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ for $a,b\in\mathbb{R}^n$, the fact that $c_1 \leq c_\ell$, and the fact that $\theta,\chi\in(0,1)$, we first have that

$$\begin{aligned}\chi\delta_k &= \chi\left(\left[\mathcal{L}^{\theta}_{c_\ell}(z_1;p_1) - \mathcal{L}^{\theta}_{c_\ell}(z_k;p_k)\right] + \left[\frac{1}{16\chi^2 c_\ell}\|\Delta^p_1\|^2\right] + \frac{a_\theta}{2\chi c_\ell}\|p_k\|^2\right) \\ &\leq \chi\left[\phi(z_0) + \frac{\zeta}{c_\ell} + \frac{a_\theta}{2\chi c_\ell}\|p_0\|^2 - \phi(z_k) + \frac{(1-\theta)^2}{2c_\ell}\|p_k\|^2\right] + \\ &\qquad \frac{1}{8\chi c_\ell}\left[\|p_1\|^2 + \|p_0\|^2\right] + \frac{1-\theta}{2c_\ell}\|p_k\|^2 \\ &\leq \chi\left[\overline{\phi} - \phi_* + \frac{\zeta}{c_\ell} + \frac{1}{2\chi c_\ell}\|p_0\|^2\right] + \frac{1}{4\chi}\left[\frac{\zeta}{c_\ell} + \frac{1}{2c_\ell}\|p_0\|^2\right] + \frac{1-\theta}{c_\ell}\|p_k\|^2 \\ &\leq \left(\chi + \frac{1}{4\chi}\right)B_\delta + \frac{1-\theta}{c_\ell}\|p_k\|^2 \\ &\leq \frac{2}{\chi}B_\delta + \frac{1-\theta}{c_\ell}\|p_k\|^2\end{aligned}$$

which is the first desired bound. To show the other bounds, we use the relation $\|z\|_1 \leq \sqrt{n}\|z\|_2$ for every $z\in\mathbb{R}^n$, the definition of $\beta_\lambda(\cdot)$, and Lemma 3.4 to obtain

$$\begin{aligned}\xi_k &= \frac{\sum_{i=2}^k c_\ell\|A^*B\Delta^y_i\|}{k-1} \leq \left(\frac{1}{k-1}\sum_{i=2}^k c_\ell^2\|A^*B\Delta^y_i\|^2\right)^{1/2} \\ &\leq \left(\frac{4c_\ell\|A\|^2}{k-1}\left[\frac{c_\ell}{4}\sum_{i=2}^k\|B\Delta^y_i\|^2\right]\right)^{1/2} \leq \left(\frac{4c_\ell\|A\|^2\delta_k}{k-1}\right)^{1/2} \leq \left[\frac{4\beta_\lambda(c_\ell)\delta_k}{k-1}\right]^{1/2},\end{aligned}$$

which implies the desired bound on ξ_k. Similarly, we use the relations $\|a\|_1 \leq \sqrt{n}\|a\|_2$ and $\|a+b+c\|^2 \leq 3\|a\|^2 + 3\|b\|^2 + 3\|c\|^2$ for every $a,b,c\in\mathbb{R}^n$, the definition of $\beta_\lambda(\cdot)$, and Lemma 3.4 to conclude that

$$
\begin{aligned}
\frac{1}{\sqrt{3}} S_{v,k} = \frac{\sum_{i=2}^{k}\|v_i\|}{\sqrt{3}(k-1)} &\le \left[\frac{1}{3(k-1)}\sum_{i=2}^{k}\|v_i\|^2\right]^{1/2} \\
&\le \frac{1}{(k-1)^{1/2}}\left[\frac{1}{\lambda^2}\sum_{i=2}^{k}\|\Delta_i^x\|^2 + \left(\frac{1}{\lambda}+M_{xy}\right)^2\sum_{i=2}^{k}\|\Delta_i^y\|^2 + \sum_{i=2}^{k} c_\ell^2\|A^*B\Delta_i^y\|^2\right]^{1/2} \\
&\le \frac{1}{(k-1)^{1/2}}\left[\left(\frac{1}{\lambda}+M_{xy}\right)^2\sum_{i=2}^{k}\left(\|\Delta_i^x\|^2+\|\Delta_i^y\|^2\right) + \left(4c_\ell\|A\|^2\right)\frac{c_\ell}{4}\sum_{i=2}^{k}\|B\Delta_i^y\|^2\right]^{1/2} \\
&\le \left[\frac{4\beta_\lambda(c_\ell)}{k-1}\right]^{1/2}\left[\frac{\alpha_\lambda}{\lambda}\sum_{i=2}^{k}\left(\|\Delta_i^x\|^2+\|\Delta_i^y\|^2\right) + \frac{c_\ell}{4}\sum_{i=2}^{k}\left(\|A\Delta_i^x\|^2+\|B\Delta_i^y\|^2\right)\right]^{1/2} \\
&\le \left[\frac{4\beta_\lambda(c_\ell)\delta_k}{k-1}\right]^{1/2}, \qquad (33)
\end{aligned}
$$

which implies the desired bound on $S_{v,k}$.

Before continuing, we briefly outline some approaches that use the results obtained so far. First, if we can show that $\|p_k\|^2/c_\ell = \mathcal{O}(1)$ for every $k \ge 1$, then (32), Lemma 3.1(a), and the fact that $\beta_\lambda(c_\ell) = \Theta(c_\ell)$ imply

$$
\min_{2\le i\le k}\|v_i\|^2 \le S_{v,k} = \mathcal{O}\left(\frac{c_\ell}{k}\right), \quad \|f_j\| = \mathcal{O}\left(\frac{1}{c_\ell}\right) \quad \forall j \ge 1.
$$

On the other hand, if we only can show that $S_{p,k} = \mathcal{O}(1)$ for every $k \ge \underline{k} \in \mathbb{N}$, then we could use a weighted sum of $\{S_{v,i}\}_{i=2}^{k}$ and $\{S_{f,i}\}_{i=2}^{k}$ — and analogous arguments as above — to bound $\|v_i\|$ and $\|f_i\|$. Our analyses in the next subsections will follow the latter approach. □

3.3 Bounding the Lagrange Multipliers

This subsection presents some specialized Lagrange multiplier bounds and generalizes the analysis in [17].

The first result presents some important relationships between p_{i-1}, p_i, and $\tilde{p}_i$.

Lemma 3.7. *The following statements hold for every $i \le k$:*

(a) $p_i = \chi\tilde{p}_i + (1-\chi)(1-\theta)p_{i-1}$;

(c) it holds that

$$
\frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota\sigma_Q^+\|\tilde{p}_i\| \le \left(\frac{1-\theta}{c_\ell}\right)\langle\tilde{p}_i, p_{i-1}\rangle + 2c_\ell\|A^*B\Delta_i^y\|D_z + 2\kappa_1
$$

where κ_1 and D_z are as in (17) and (15), respectively.

Proof. (a) This is an immediate consequence of the updates for p_i and $\tilde{p}_i$ in the DP.ADMM.

(b) Let $i \le k$ be fixed and define the auxiliary residual

$$
\tilde{v}_i := v_i - \nabla f(z_i) - Q^*\tilde{p}_i.
$$

Using the fact that $\tilde{p}_i \in Q(\mathbb{R}^n)$, Lemma A.1 with $(S,u) = (Q, \tilde{p}_i)$, Lemma 3.1(c), and the triangle inequality, we have that

$$
\begin{aligned}
\frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota \sigma_Q^+ \|\tilde{p}_i\| &\le \frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota \|Q^*\tilde{p}_i\| = \frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota \|v_i - \nabla f(z_i) - \tilde{v}_i\| \\
&\le \frac{1}{c_\ell}\|\tilde{p}_k\|^2 + d_\iota \left[\|\tilde{v}_i\| + \|\nabla f(z_i)\| + \|v_i\|\right] \\
&\le \frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota \left[\|\tilde{v}_i\| + (M_{xy} + \frac{1}{\lambda})D_z + G_f + c_\ell \|A^* B \Delta_i^y\|\right] \\
&\le \frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota \|\tilde{v}_i\| + c_\ell \|A^* B \Delta_i^y\| D_z + \kappa_1 - K_h D_z. \qquad (34)
\end{aligned}
$$

We now derive a suitable bound on $d_\iota \|\tilde{v}_i\|$. First, note that Lemma 3.1(c) and the definition of $\tilde{v}_i$ imply that $\tilde{v}_i \in \partial h(z_i)$. Using the definition of D_z in (15) and Lemma A.3 with $(\psi, r, z, \bar{z}) = (h, \tilde{v}_i, z_i, z_\iota)$, it now follows that

$$
\begin{aligned}
d_\iota \|\tilde{v}_i\| = \|\tilde{v}_i\| \mathrm{dist}_{\partial Z}(z_\iota) &\le \left[\mathrm{dist}_{\partial Z}(z_\iota) + \|z_i - z_\iota\|\right] K_h + \langle \tilde{v}_i, z_i - z_\iota \rangle \\
&\le 2K_h D_z + \langle \tilde{v}_i, z_i - z_\iota \rangle. \qquad (35)
\end{aligned}
$$

On the other hand, Lemma 3.1(c) and the definitions of κ_1 and $\tilde{p}_i$ imply that

$$
\begin{aligned}
&\langle \tilde{v}_i, z_i - z_\iota \rangle \\
&= \langle v_i - \nabla f(z_i) - Q^* \tilde{p}_i, z_i - z_\iota \rangle \\
&\le (\|v_i\| + \|\nabla f(z_i)\|) \|z_i - z_\iota\| - \langle \tilde{p}_i, Q z_i - d \rangle \\
&\le \left[\left(M_{xy} + \frac{1}{\lambda}\right) D_z + G_f + c_\ell \|A^* B \Delta_i^y\|\right] D_z - \langle \tilde{p}_i, Q z_i - d \rangle \\
&= \kappa_1 - K_h D_z + c_\ell \|A^* B \Delta_i^y\| D_z + \left(\frac{1-\theta}{c}\right) \langle \tilde{p}_i, p_{i-1} \rangle - \frac{1}{c}\|\tilde{p}_i\|^2. \qquad (36)
\end{aligned}
$$

The conclusion now follow from combining (34), (35), and (36). □

The next result establishes bounds on $\|p_i\|$ and $S_{p,k}$.

Lemma 3.8. *Let κ_2 and $\tau_1(\cdot)$ be as in (17) and (18), respectively. Then, the following statements hold:*

(a) for every $i \le k$, it holds that $\|p_i\| \le \tau_1(c_\ell)$.

(b) if $k \ge 2$, it holds that

$$
S_{p,k} \le \kappa_2 + \left[\frac{4\chi D_z}{(\chi + \theta - \chi\theta) d_\iota \sigma_Q^+}\right] \xi_k.
$$

Proof. (a) We proceed by induction on i. The case for $i = 0$ is immediate from the definition of $\tau_1(\cdot)$ in (18). Suppose now that the bound holds for some $i \ge 0$. Using Lemma 3.7(b) and the

Cauchy-Schwarz inequality, we have that

$$
\begin{aligned}
&\left(d_\iota + \frac{1}{c_\ell \sigma_Q^+}\|\tilde{p}_{i+1}\|\right)\|\tilde{p}_{i+1}\| = \frac{1}{\sigma_Q^+}\left[\frac{1}{c_\ell}\|\tilde{p}_{i+1}\|^2 + d_\iota \sigma_Q^+ \|\tilde{p}_{i+1}\|\right] \\
&\le \frac{1}{\sigma_Q^+}\left[\left(\frac{1-\theta}{c_\ell}\right)\langle \tilde{p}_i, p_{i-1}\rangle + 2c_\ell \|A^* B \Delta_i^y\| D_z + 2\kappa_1\right] \\
&\le \frac{1}{\sigma_Q^+}\left[\frac{1}{c_\ell}\|\tilde{p}_{i+1}\| \cdot \|p_i\| + 2c_\ell \|A^* B\| D_z^2 + 2\kappa_1\right] \\
&= 2d_\iota \left(\frac{\kappa_1 + c_\ell \|A^* B\| D_z^2}{d_\iota \sigma_Q^+}\right) + \left(\frac{1}{c_\ell \sigma_Q^+}\|\tilde{p}_{i+1}\|\right)\|p_i\| \le \left(d_\iota + \frac{1}{c_\ell \sigma_Q^+}\|\tilde{p}_{i+1}\|\right)\tau_1(c_\ell),
\end{aligned}
$$

where the last inequality follows from the induction hypothesis and the definition of $\tau_1(\cdot)$. Hence, we have $\|\tilde{p}_{k+1}\| \le \tau_1(\cdot)$. Using the previous bound, Lemma 3.7(a), and the induction hypothesis, we conclude that

$$
\begin{aligned}
\|p_{i+1}\| &\le \chi \|\tilde{p}_{i+1}\| + (1-\chi)(1-\theta)\|p_i\| \\
&\le \chi \tau_1(c_\ell) + (1-\chi)(1-\theta)\tau_1(c_\ell) \le \tau_1(c_\ell).
\end{aligned}
$$

(b) Let $k \ge 2$ and $i \le k$ be arbitrary, and define

$$
\nu_i(c_\ell) := \kappa_1 + c_\ell \|A^* B \Delta_i^y\| D_z, \quad e_0 := (1-\theta)(1-\chi).
$$

Using Lemma 3.7(a) thrice, and the bounds $2ab \le a^2 + b^2$ and $(a+b)^2 \le 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}_+$, we first have that

$$
\begin{aligned}
&\frac{1}{c_\ell}\|p_i\|^2 + d_\iota \sigma_Q^+ \|p_i\| \\
&= \frac{1}{c_\ell}\|\chi \tilde{p}_i + e_0 p_{i-1}\|^2 + d_\iota \sigma_Q^+ \|\chi \tilde{p}_i + e_0 p_{i-1}\| \\
&\le 2\chi \left[\frac{1}{c_\ell}\|\tilde{p}_i\|^2 + d_\iota \sigma_Q^+ \|\tilde{p}_i\|\right] + \frac{2e_0^2}{c_\ell}\|p_{i-1}\|^2 + e_0 d_\iota \sigma_Q^+ \|p_{i-1}\| \\
&\le 2\chi \left[\frac{1-\theta}{c_\ell}\langle \tilde{p}_i, p_{i-1}\rangle + 2\nu_i(c_\ell)\right] + \frac{2e_0^2}{c_\ell}\|p_{i-1}\|^2 + e_0 d_\iota \sigma_Q^+ \|p_{i-1}\| \\
&= \frac{2(1-\theta)}{c_\ell}\langle p_i, p_{i-1}\rangle + \underbrace{\frac{2e_0(e_0-1)}{c_\ell}}_{\le 0}\|p_{i-1}\|^2 + e_0 d_\iota \sigma_Q^+ \|p_{i-1}\| + 4\chi \nu_i(c_\ell) \\
&\le \frac{1-\theta}{c_\ell}\|p_i\|^2 + \frac{1-\theta}{c_\ell}\|p_{i-1}\|^2 + e_0 d_\iota \sigma_Q^+ \|p_{i-1}\| + 4\chi \nu_i(c_\ell).
\end{aligned} \tag{37}
$$

Rearranging terms yields

$$
\begin{aligned}
&\frac{2\theta - 1}{c_\ell}\|p_i\|^2 + (1-e_0) d_\iota \sigma_Q^+ \|p_i\| \\
&\le \frac{1-\theta}{c_\ell}\left[\|p_{i-1}\|^2 - \|p_i\|^2\right] + e_0 d_\iota \sigma_Q^+ \left[\|p_{i-1}\| - \|p_i\|\right] + 4\chi \nu_i(c_\ell)
\end{aligned}
$$

Summing the above inequality from $i = 2$ to k and combining the resulting bound with Lemma 3.5(a), (16), and the definition of κ_2 in (17), we conclude that

$$
\begin{aligned}
(1-e_0)d_\iota\sigma_Q^+\sum_{i=2}^{k}\|p_i\| &\le \frac{2\theta-1}{c_\ell}\sum_{i=2}^{k}\|p_i\|^2 + (1-e_0)d_\iota\sigma_Q^+\sum_{i=2}^{k}\|p_i\| \\
&\le \frac{1-\theta}{c_\ell}\left[\|p_1\|^2 - \|p_k\|^2\right] + e_0 d_\iota\sigma_Q^+\left[\|p_1\| - \|p_k\|\right] + 4\chi\sum_{i=2}^{k}\nu_i(c_\ell) \\
&\le \frac{1}{c_1}\|p_1\|^2 + e_0 d_\iota\sigma_Q^+\|p_1\| + 4\chi(k-1)(D_z\xi_k + \kappa_1) \\
&\le 2\zeta c_1^{-1} + \sqrt{2\zeta}e_0 d_\iota\sigma_Q^+ + \|p_0\|^2 + 4\chi(k-1)(D_z\xi_k + \kappa_1) \\
&\le (k-1)\left[(1-e_0)d_\iota\sigma_Q^+\kappa_2 + 4\chi D_z\xi_k\right],
\end{aligned}
$$

which, in view of the definitions of e_0 and $S_{p,k}$, implies the desired bound. □

3.4 Proof of Proposition 2.1

This subsection gives the proof of Proposition 2.1 and follows the setting described at the end of Subsection 3.2.

We first show that several quantities become nicely bounded for sufficiently large enough k.

Lemma 3.9. *Let κ_3 and $\tau_3(\cdot)$ be as in (17) and (18), respectively. Then, the following statements hold:*

(a) for every $k \ge 1 + \xi^{-2}\tau_3(c_\ell)$, it holds that $\xi_k \le \xi$;

(b) if $\xi_k \le \xi$, then $S_{p,k} \le \kappa_3$.

Proof. (a) Let $k \ge 1 + \xi^{-2}\tau_3(c_\ell)$. Using Lemmas 3.6 and 3.8(a), it holds that

$$
\xi_k \le \sqrt{\frac{12\beta_\lambda(c_\ell)\cdot\delta_k}{k-1}} \le \xi\sqrt{\frac{12\chi^{-1}\beta_\lambda(c_\ell)\left[2\chi^{-1}B_\delta + c_\ell^{-1}(1-\theta)\|p_k\|^2\right]}{\tau_3(c_\ell)}} \le \xi.
$$

(b) This is immediate from Lemma 3.8(b), the definition of κ_3, and our assumption on ξ_k. □

We next show some important bounds involving $S_{v,j}$, $S_{f,j}$, and r_k.

Lemma 3.10. *Let (κ_3, κ_4) and (τ_2, τ_4) be as in (17) and (18), respectively. If $\xi_k \le \xi$ and $k \ge 2$, then the following statements hold*

(a) $\sum_{j=2}^{k}(j-1)S_{v,j} \le (k-1)^{3/2}\kappa_4\sqrt{\tau_2(c_\ell)}$;

(b) $\sum_{j=2}^{k}(j-1)S_{f,j} \le 2\,(\zeta+\kappa_3)\,(k-1)^2/(\chi c_\ell)$;

(c) $r_k \le (k-1)^{-1/2}\tau_4(c_\ell)$.

Proof. (a) Let δ_j be as in Lemma 3.4. Using the definition of $\tau_2(c_\ell)$, the fact that $\xi_k \le \xi$ and $c_\ell \ge c_1$, Lemmas 3.9 and 3.6, and the bound $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for $a, b \in \mathbb{R}_+$, it follows that

$$\begin{aligned}
\sum_{j=2}^{k}(j-1)S_{v,j} &= \sum_{j=2}^{k}\sum_{i=2}^{j}\|v_i\| \le \sum_{j=2}^{k}\sqrt{12(j-1)\beta_\lambda(c_\ell)\cdot\delta_j} \\
&\le \sum_{j=2}^{k}\sqrt{12(j-1)\chi^{-1}\beta_\lambda(c_\ell)\left[2\chi^{-1}B_\delta + c_\ell^{-1}(1-\theta)\|p_j\|^2\right]} \\
&\le \sqrt{\tau_2(c_\ell)}\left[(k-1)^{3/2}\sqrt{\frac{2B_\delta}{\chi}} + (k-1)^{1/2}\sqrt{\frac{(1-\theta)}{c_1}}\sum_{j=2}^{k}\|p_j\|\right] \\
&\le (k-1)^{3/2}\sqrt{\tau_2(c_\ell)}\left(\sqrt{\frac{2B_\delta}{\chi}} + \kappa_3\sqrt{\frac{1-\theta}{c_1}}\right) \\
&= (k-1)^{3/2}\kappa_4\sqrt{\tau_2(c_\ell)}.
\end{aligned}$$

(b) Let δ_j be as in part (a). Using Lemmas 3.8(b), Lemma 3.1, Lemma 3.5(a), the triangle inequality, and the fact that $\theta \in (0,1)$ and $k \ge 2$, it follows that

$$\begin{aligned}
\sum_{j=2}^{k}(j-1)S_{f,j} &= \sum_{j=2}^{k}\sum_{i=2}^{j}\|f_i\| = \frac{1}{\chi c_\ell}\sum_{j=2}^{k}\sum_{i=2}^{j}\|p_i - (1-\theta)p_{i-1}\| \\
&\le \frac{1}{\chi c_\ell}\sum_{j=2}^{k}\left(\|p_1\| + 2\sum_{i=2}^{j}\|p_i\|\right) \le \frac{2}{\chi c_\ell}(\zeta + \kappa_3)\sum_{j=2}^{k}(j-1) \\
&\le \frac{2}{\chi c_\ell}(\zeta + \kappa_3)(k-1)^2
\end{aligned}$$

(c) Combining parts (a) and (b), we have that

$$\begin{aligned}
\frac{(k-1)^2}{2}r_k &\le \frac{(k-1)^2}{2}\min_{2\le j\le k}\left(S_{v,j} + \sqrt{\frac{c_\ell^3}{k-1}}S_{f,j}\right) = \sum_{j=2}^{k}(j-1)\left(S_{v,j} + \sqrt{\frac{c_\ell^3}{k-1}}S_{f,j}\right) \\
&\le (k-1)^{3/2}\left[\kappa_4\sqrt{\tau_2(c_\ell)} + \frac{2\sqrt{c_\ell}}{\chi}(\zeta + \kappa_3)\right] = \frac{(k-1)^{3/2}\tau_4(c_\ell)}{2},
\end{aligned}$$

which implies the desired bound. □

We are now ready to give the proof of Proposition 2.1.

Proof of Proposition 2.1. (a) This is immediate from the termination condition in step 2b and the inclusion in Lemma 3.1(c).

(b) Let $\ell \ge 1$ be an arbitrary cycle index and $\overline{T}_\ell(\rho)$ be as in (19). Suppose for contradiction that the current cycle has not terminated by the beginning of iteration $\overline{T}_\ell(\rho)$. Using Lemma 3.9(a), it holds that $\xi_k \le \xi$, and hence, using Lemma 3.10(c), it follows that

$$r_{\overline{T}_\ell(\rho)} \le \frac{\tau_4(c_\ell)}{\sqrt{\overline{T}_\ell(\rho) - 1}} \le \frac{\tau_4(c_\ell)}{\sqrt{\tau_4^2(c_\ell)\rho^{-2}}} = \rho,$$

which contradicts our assumption that the cycle has not terminated due to step 2c of the DP.ADMM. Hence, it must be the case that each cycle contains at most $\overline{T}_\ell(\rho)$ iterations.

(c) Let $\overline{c}(\eta)$ be as in (20) and $\ell \geq 1$ be a cycle index where $c_\ell \geq \overline{c}(\eta)$. In view of part (a), let $k \leq \overline{T}_\ell(\rho)$ be the first index where $r_k \leq \rho$ and let $j \leq k$ be the first index where $r_k = \|v_j\| + (k-1)^{-1/2} c_\ell^{3/2} \|f_j\|$. Using step 2c of the DP.ADMM, Lemma 3.10(c), the fact that $c_\ell \geq \overline{c}(\eta) \geq 1$, and the bounds $\tau_2(c_\ell) \leq c_\ell \cdot \tau_2(1)$, we have that $\|v_j\| \leq r_k \leq \rho$ and

$$\begin{aligned}
\|f_j\| &\leq \sqrt{\frac{k-1}{c_\ell^3}} \left[\|v_j\| + \frac{c_\ell^{3/2}\|f_j\|}{\sqrt{k-1}}\right] \leq \frac{\tau_4(c_\ell)}{c_\ell^{3/2}} = \frac{1}{2c_\ell^{3/2}} \left[\kappa_4 \sqrt{\tau_2(c_\ell)} + \frac{2\sqrt{c_\ell}}{\chi}(\zeta + \kappa_3)\right] \\
&\leq \frac{1}{2c_\ell^{3/2}} \left[\sqrt{c_\ell}\left(\kappa_4\sqrt{\tau_2(1)} + \frac{2}{\chi}\left[\zeta + \kappa_3\right]\right)\right] = \frac{1}{2c_\ell}\left[\kappa_4\sqrt{\tau_2(1)} + \frac{2}{\chi}(\zeta+\kappa_3)\right] = \frac{\eta\overline{c}(\eta)}{c_\ell} \leq \eta.
\end{aligned}$$

In view of steps 2b–2c of the DP.ADMM, it follows that the method must first stop in step 2b at iteration j. $\square$

4 Concluding Remarks

Under a careful examination of the proofs of Section 3, we can see that assumption (A2) is only used to provide a conservative bound on the quantity $\phi(z_0) - \phi(z_k)$ in the proof of Lemma 3.6. Hence, (A2) may be relaxed to requiring $\phi(z_0) - \phi(z_k)$ be bounded for every $k \geq 1$ and every cycle. One particular relaxation is the case where $(\bar{x}_0, \bar{y}_0) \in \mathcal{F}$ and $z_0 = (\bar{x}_0, \bar{y}_0)$ for every cycle (cf. the remarks about Algorithm 2.1 in Subsection 2.2). In this case, one can remove the assumption that $\overline{\phi} < \infty$ as $z_0 \in Z$ for every cycle.

A Technical Inequalities

The first result, whose proof can be found in [12, Lemma 1.3], presents a relationship between elements in the image of a linear operator.

Lemma A.1. *For any $S \in \mathbb{R}^{m\times n}$ and $u \in S(\mathbb{R}^{m\times n})$, we have $\sigma_S^+\|u\| \leq \|Su\|$.*

The proof of the following result can be found in [17, Lemma B.2].

Lemma A.2. *For any $(\tau, \theta) \in [0,1]^2$ satisfying $\tau \leq \theta^2$ and any $a, b \in \mathbb{R}^n$, we have that*

$$\|a - (1-\theta)b\|^2 - \tau\|a\|^2 \geq \left[\frac{(1-\tau)-(1-\theta)^2}{2}\right]\left(\|a\|^2 - \|b\|^2\right). \tag{38}$$

Finally, the proof of the next result can be found in [18, Lemma 4.7].

Lemma A.3. *Suppose $\psi \in \overline{\text{Conv}}\ \mathbb{R}^n$ is K_ψ-Lipschitz continuous. Then, for every $z, \bar{z} \in \operatorname{dom}\psi$ and $r \in \partial\psi(z)$, it holds that*

$$\|r\|\text{dist}_{\partial(\operatorname{dom}\psi)}(\bar{z}) \leq \left[\text{dist}_{\partial(\operatorname{dom}\psi)}(\bar{z}) + \|z - \bar{z}\|\right] K_\psi + \langle r, z - \bar{z}\rangle.$$

B Implementation Details

This appendix presents algorithms for implementing the more involved parts of the DP.ADMM. It contains two sub-appendices. The first one presents an iterative method for step 0 of the DP.ADMM whereas the second one presents a dynamic algorithm for step 2c of the DP.ADMM.

B.1 Implementing Step 0 of the DP.ADMM

Recall that the goal of step 0 of the DP.ADMM is to obtain a pair $(x_0, y_0) \in Z$ satisfying $c_\ell \Phi_\ell(x_0, y_0) \le \zeta$. In this sub-appendix, we show that such a pair can be obtained by applying an accelerated composite gradient (ACG) algorithm to the composite problem $\min_{(x,y)\in Z} c_\ell \Phi_\ell(x, y)$ for $\mathcal{O}(c_\ell)$ ACG iterations.

We first review the (FISTA) ACG variant studied in [15, 16]. Its problem of interest is

$$\min_{z\in\mathbb{R}^n} \{\psi(z) := \psi_s(z) + \psi_n(z)\} \tag{39}$$

where $\psi_n \in \overline{\text{Conv}}\,(Z)$, ψ_s is continuously differentiable on Z, and $\psi_s(u) - \ell_{\psi_s}(u; z) \le L\|u - z\|^2/2$ for every $u, z \in Z$. Given an initial point z_0 and a restricted stepsize $\lambda > 0$, we state the full algorithm in Algorithm B.1.

Algorithm B.1: Accelerated Composite Gradient (ACG) Algorithm

Require: $z_0 \in Z$, $\lambda \in (0, 1/L]$, and $\sigma \in (0, 1]$.

Function `ACG(`$[\psi_s, \psi_n], L, z_0, \sigma$`)`**:**
- `STEP 0` (initialization):
- Set $A_0 \leftarrow 0$, $\Gamma_0 \leftarrow 0$, $z_0^c \leftarrow z_0$
- **for** $j = 1, \ldots$ **do**
  - `STEP 1` (main iterates):
  - $a_{j-1} \leftarrow \frac{\lambda + \sqrt{\lambda^2 + 4\lambda A_{j-1}}}{2}$
  - $A_j \leftarrow A_{j-1} + a_{j-1}$
  - $\tilde{z}_{j-1} \leftarrow \frac{A_{j-1}}{A_j} z_{j-1} + \frac{a_{j-1}}{A_j} z_{j-1}^c$
  - $z_j \leftarrow \text{argmin}_{u\in\mathbb{R}^n} \left\{\lambda \left[\ell_{\psi_s}(u; \tilde{z}_{j-1}) + \psi_n(u)\right] + \frac{1}{2}\|u - \tilde{z}_{j-1}\|^2\right\}$
  - $z_j^c \leftarrow z_{j-1}^c + \frac{a_{j-1}}{\lambda}(z_j - \tilde{z}_{j-1})$
  - `STEP 2` (auxiliary iterates):
  - $\gamma_{j-1}(\cdot) \leftarrow \ell_{\psi_s}(z_j; \tilde{z}_{j-1}) + \psi_n(z_j) + \frac{1}{\lambda}\langle \tilde{z}_{j-1} - z_j, \cdot - z_j\rangle$
  - $\Gamma_j(\cdot) \leftarrow \frac{A_{j-1}}{A_j}\Gamma_{j-1}(\cdot) + \frac{a_{j-1}}{A_j}\gamma_{j-1}(\cdot)$
  - $\tilde{v}_j \leftarrow \frac{z_0^c - z_j^c}{A_j}$
  - $\tilde{\varepsilon}_j \leftarrow (\psi_s + \psi_n)(z_j) - \Gamma_j(z_j^c) - \left\langle \tilde{v}_j, z_j - z_j^c\right\rangle$
  - `STEP 3` (termination check):
  - **if** $\|\tilde{v}_j\|^2 + 2\tilde{\varepsilon}_j \le \sigma^2\|z_j - z_0\|^2$ **then**
    - **return** $(z_j, \tilde{v}_j, \tilde{\varepsilon}_j)$

The result below presents some basic properties about the ACG algorithm.

Lemma B.1. *The following properties hold about the ACG algorithm in Algorithm B.1:*

(a) for every $j \ge 1$, it holds that $\tilde{v}_j \in \partial_{\tilde{\varepsilon}_j}(\psi_s + \psi_n)(z_j)$ and

$$\|A_j\tilde{v}_j + z_j - z_0\|^2 + 2A_j\tilde{\varepsilon}_j \le \|z_j - z_0\|^2, \quad A_j \ge \frac{\lambda(j-1)^2}{4}; \tag{40}$$

(b) it stops in at most $\left\lceil 1 + (2\sqrt{2})/(\sigma\sqrt{\lambda})\right\rceil$ ACG iterations.

Proof. (a) See [15, Proposition 2.2.3] with $\mu = 0$ and $\lambda_k = \lambda$ for every $k \geq 1$.

(b) Let $\left\lceil 1 + (2\sqrt{2})/(\sigma\sqrt{\lambda}) \right\rceil$ and observe that the second inequality in (40) implies that $A_j \geq \lambda(j-1)^2/4 \geq 2/\sigma^2$. Using the previous bound, the first inequality in (40), the relation $(a+b)^2 \leq 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}^2$, and the fact that $\sigma \in (0,1]$ implies $A_j \geq 2$, it holds that

$$\begin{aligned}
\|\tilde{v}_j\|^2 + 2\tilde{\varepsilon}_j &\leq \max\left\{1/A_j^2, 1/(2A_j)\right\}\left(\|A\tilde{v}_j\|^2 + 4A_j\tilde{\varepsilon}_j\right) \\
&\leq \max\left\{1/A_j^2, 1/(2A_j)\right\}\left(2\|A\tilde{v}_j + z_j - z_0\|^2 + 2\|z_j - z_0\|^2 + 4A_j\tilde{\varepsilon}_j\right) \\
&\leq \max\left\{(2/A_j)^2, 2/A_j\right\}\|z_j - z_0\|^2 \leq \sigma^2\|z_j - z_0\|^2,
\end{aligned}$$

and hence the ACG must stop at or before iteration j. □

We now make a few remarks about the above properties. First, if $(z_j, \tilde{v}_j, \tilde{\varepsilon}_j)$ is obtained with $\sigma = 0$ then z_j is a global minimum of (39). Hence, the parameter σ can be viewed as a type of tolerance parameter for the ACG. Second, using the relation $(a+b)^2 \leq 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}$, it is straightforward to show that

$$\|\tilde{v}_j\|^2 \leq \frac{2\|A_j\tilde{v}_j + z_j - z_0\|^2 + 2\|z_j - z_0\|^2}{A_j^2} \leq \frac{4D_z^2}{A_j^2}, \quad \tilde{\varepsilon}_j \leq \frac{D_z^2}{2A_j}.$$

Moreover, since $\lim_{j\to\infty} A_j \to \infty$, we have that $\lim_{j\to\infty} \max\{\|\tilde{v}_j\|, \tilde{\varepsilon}_j\} \to 0$ and hence that z_j converges to a global minimum of (39).

We now describe how to call the ACG in Algorithm B.2 to implement step 0 of the DP.ADMM. Using the fact that $\Phi_\ell(\cdot)$ is continuously differentiable on Z and its gradient is $c_\ell\|Q\|^2$-Lipschitz continuous, and hence $\Phi_\ell(u) - \ell_{\Phi_\ell}(u; z) \leq (c_\ell\|Q\|^2)\|u - z\|^2/2$ for every $u, z \in Z$, we give the explicit call in Algorithm B.2.

Algorithm B.2: Initializing a cycle

Require: $\bar{z}_{\ell-1} \in Z$

Function `CycleInit`$(\bar{z}_{\ell-1}, c_\ell, Q, d)$:
- $(\psi_s, \psi_n) \leftarrow (c_\ell\Phi_\ell, \delta_Z)$
- $(z_0, \tilde{v}, \tilde{\varepsilon}) \leftarrow$ `ACG`$([\psi_s, \psi_n], 1/(c_\ell\|Q\|)^2, \bar{z}_{\ell-1}, 1/\sqrt{2})$
- **return** z_0

The lemma below describes the key properties of Algorithm B.2.

Lemma B.2. *The following properties hold about Algorithm B.2.*

(a) it stops in at most $\lceil 1 + 8c_\ell\|Q\| \rceil$ ACG iterations;

(b) its output z_0 satisfies

$$c_\ell\Phi_\ell(z_0) \leq \frac{(1-\theta)^2}{2}\|p_0\|^2 + 2D_z^2.$$

Proof. (a) This is an immediate consequence of Lemma B.1(b) and the chosen inputs of the ACG call.

(a) It follows from part (a), Lemma B.1(a), and step 3 of the ACG algorithm that

$$\tilde{v} \in \partial_{\tilde{\varepsilon}}(c_\ell\psi_\ell)(\bar{z}_{\ell-1}), \quad \|\tilde{v}\|^2 + 2\tilde{\varepsilon} \leq \sigma^2\|z_0 - \bar{z}_{\ell-1}\|^2.$$

In particular, the above inclusion, the relation $2\langle a,b\rangle \geq -\|a\|^2 - \|b\|^2$ for every $a,b\in\mathbb{R}^n$, and the fact that $\sigma \leq 1$, imply that for any $u\in\mathcal{F}$, we have

$$\begin{aligned}\frac{(1-\theta)^2}{2}\|p_0\|^2 = c_\ell\Phi_\ell(u) &\geq c_\ell\Phi_\ell(z_0) + \langle \tilde{v}, u-\bar{z}_{\ell-1}\rangle - \tilde{\varepsilon} \\ &\geq c_\ell\Phi_\ell(z_0) - \frac{1}{2}\left(\|\tilde{v}\|^2 + 2\|u-\bar{z}_{\ell-1}\|^2 + 2\tilde{\varepsilon}\right) \\ &\geq c_\ell\Phi_\ell(z_0) - \frac{1}{2}\left(\sigma^2\|z_0-\bar{z}_{\ell-1}\|^2 + 2D_z^2\right) \geq c_\ell\Phi_\ell(z_0) - 2D_z^2,\end{aligned}$$

which clearly implies the desired bound. □

From the above results, it is clear that: (i) the output z_0 satisfies the requirements of step 0 of the DP.ADMM with $\zeta = (1-\theta)^2\|p_0\|^2/2 + 2D_z^2$; and (ii) obtaining z_0 requires $\mathcal{O}(c_\ell)$ ACG iterations.

B.2 Implementing Step 2c of the DP.ADMM

Recall that in step 2c of the DP.ADMM, the computation of r_k at cycle iteration k requires evaluating the minimum of k scalars that change iteration to iteration. In this sub-appendix, we present a dynamic algorithm that computes r_k in both $\mathcal{O}(1)$ storage and $\mathcal{O}(1)$ runtime.

We first remark that indices j for which $r_j \leq \rho$ holds all have the property that either (i) $\|v_j\| < \rho$ or (ii) $\|v_j\| = \rho$ and $\|f_j\| = 0$. Since the latter condition immediately implies that $r_j \leq \rho$, it suffices to consider an algorithm to determine when the former condition implies $r_j \leq \rho$. Using the definition of r_j and the fact that the function $k \mapsto c_\ell^{3/2}(k-1)^{-1/2}\|f_j\|$ is strictly decreasing for fixed ℓ and j, it is straightforward to see that there is a threshold index k_j for which $\|v_j\| + c_\ell^{3/2}(k_j-1)^{-1/2}\|f_j\| \leq \rho$ for every $k_j \geq k$. Hence, we can consider an algorithm that dynamically computes $t_k := \min_{j\leq k} k_j$ and stops the current cycle, i.e. the stopping condition of step 2c, whenever $k \leq t_k$ and $\xi_k \leq \xi$.

For completeness, we show how to implement this dynamic algorithm in Algorithm B.3. As claimed at the beginning of this sub-appendix, this algorithm only needs $\mathcal{O}(1)$ storage and $\mathcal{O}(1)$ runtime during each cycle iteration.

Algorithm B.3: An efficient cycle termination check

$k_{\max} \leftarrow \infty$
if $k \geq 2$ **then**
  $\xi_k \leftarrow \sum_{i=2}^{k} c_\ell \|A^* B \Delta_i^y\|/(k-1)$
  if $k \geq k_{\max}$ **or** $(\|v_k\| = \rho$ and $\|f_j\| = 0)$ **then**
    $(c_{\ell+1}, \bar{x}_\ell, \bar{y}_\ell) \leftarrow (2c_\ell, x_k, y_k)$
    break loop
  else if $\xi_k \leq \xi$ **and** $\|v_k\| < \rho$ **then**
    $\pi_k \leftarrow c_\ell^{3/2}\|f_k\|/(\rho - \|v_k\|)$
    $k_{\max} \leftarrow \min\{k_{\max}, \lceil 1 + \pi_k^2 \rceil\}$

References

[1] Dimitri P Bertsekas. *Nonlinear programming.* Taylor & Francis, 3ed edition, 2016.

[2] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[3] Mian Tao Chao, Ye Zhang, and Jin Bao Jian. An inertial proximal alternating direction method of multipliers for nonconvex optimization. *International Journal of Computer Mathematics*, pages 1–19, 2020.

[4] Jonathan Eckstein and Dimitri P Bertsekas. On the douglas–rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

[5] Jonathan Eckstein and Michael C Ferris. Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS Journal on Computing*, 10(2):218–235, 1998.

[6] Jonathan Eckstein and Masao Fukushima. Some reformulations and applications of the alternating direction method of multipliers. In *Large scale optimization*, pages 115–134. Springer, 1994.

[7] Jonathan Eckstein and Benar Fux Svaiter. A family of projective splitting methods for the sum of two maximal monotone operators. *Mathematical Programming*, 111(1):173–199, 2008.

[8] Jonathan Eckstein and Benar Fux Svaiter. General projective splitting methods for sums of maximal monotone operators. *SIAM Journal on Control and Optimization*, 48(2):787–811, 2009.

[9] Daniel Gabay. Applications of the method of multipliers to variational inequalities. In *Studies in mathematics and its applications*, volume 15, pages 299–331. Elsevier, 1983.

[10] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & mathematics with applications*, 2(1):17–40, 1976.

[11] Roland Glowinski and Americo Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.

[12] Max L. N. Goncalves, Jefferson G. Melo, and Renato D. C. Monteiro. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *arXiv preprint arXiv:1702.01850*, 2017.

[13] Zehui Jia, Jieru Huang, and Zhongming Wu. An incremental aggregated proximal ADMM for linearly constrained nonconvex optimization with application to sparse logistic regression problems. *Journal of Computational and Applied Mathematics*, 390:113384, 2021.

[14] Bo Jiang, Tianyi Lin, Shiqian Ma, and Shuzhong Zhang. Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis. *Computational Optimization and Applications*, 72(1):115–157, 2019.

[15] Weiwei Kong. Accelerated inexact first-order methods for solving nonconvex composite optimization problems. *arXiv preprint arXiv:2104.09685*, 2021.

[16] Weiwei Kong and Renato D. C. Monteiro. Accelerated inexact composite gradient methods for nonconvex spectral optimization problems. *arXiv preprint arXiv:2007.11772*, 2020.

[17] Weiwei Kong and Renato D. C. Monteiro. An accelerated inexact dampened augmented Lagrangian method for linearly-constrained nonconvex composite optimization problems. *arXiv preprint arXiv:2110.11151*, 2021.

[18] Jefferson G. Melo and Renato D. C. Monteiro. Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical lagrangian function and a full Lagrange multiplier update. *Available on arXiv:2008.00562*, 2020.

[19] Ralph Tyrrell Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research*, 1(2):97–116, 1976.

[20] Andrzej Ruszczyński. An augmented Lagrangian decomposition method for block diagonal linear programming problems. *Operations Research Letters*, 8(5):287–294, 1989.

[21] Kaizhao Sun and Andy Sun. Dual descent alm and admm. *arXiv preprint arXiv:2109.13214*, 2021.

[22] Andreas Themelis and Panagiotis Patrinos. Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

[23] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

[24] Jiawei Zhang and Zhi-Quan Luo. A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization. *SIAM Journal on Optimization*, 30(3):2272–2302, 2020.